Phase Transitions in Sparsely Coded Neural Networks

Jacob Carroll[1,3] & Nils Carlson[2,3]   Mentor: Garrett Kenyon[3]

[1]Virginia Tech [2]New Mexico Tech [3]Los Alamos National Laboratory

LA-UR-17-26726

Background

Image recognition is an extremely difficult task for computers. Humans have no trouble with this task, so why not try to mimic how people see?

We use PetaVision (a high-performance neural simulation toolbox)[1] to replicate the behaviour of the visual cortex by training a convolutional neural network to sparsely code images using a learned dictionary of features $\{\phi_m\}$.

Given an input $\mathbf{s}$, PetaVision creates a sparse reconstruction $\hat{\mathbf{s}}$ such that

$$\hat{\mathbf{s}} = \sum_m a_m \phi_m \tag{1}$$

where a_m is the activity of the m^{th} neuron, and ϕ_m is the m^{th} feature kernel. This is done by minimizing the following energy function:

$$E = \frac{1}{2}\underbrace{\|\mathbf{s}-\hat{\mathbf{s}}\|^2}_{\text{Reconstruction error}} + \lambda \underbrace{\sum |a_m|}_{\text{Activation cost}} \tag{2}$$

via a combination of gradient and stochastic gradient descent on a_m and ϕ_m respectively, where λ determines sparsity[2]. We investigate how PetaVision's behavior changes as λ varies and observe a minima in the reconstruction error of our network for a certain value of sparsity.

Minima in system parameters can indicate that a phase transition is occurring.

If this is the case in PetaVision, then we can take advantage of the scaling laws inherent to phase transitions to be able to predict the most efficient sparsity for any network size.

Phase Transitions

A phase transition occurs when a system shifts between two (or more) different regimes of dynamics (phases).

A critical point is the point "in between" these phases, where the system cannot be said to be in one phase or the other.

Critical points exist at maxima or minima of the system undergoing a phase transition, and follow very particular scaling laws as the system size is changed.

In this case the potential critical point we are investigating is the minima in percent reconstruction error, and the system size corresponds to the number of neurons in our network.

If our system is undergoing a phase transition, and the minima in percent reconstruction error is a critical point, then our system should obey the following scaling laws:

$$\text{Value at Minima} \sim L^{\gamma/\nu} \tag{3}$$

$$\text{Location of Minima} \sim L^{1/\nu} \tag{4}$$

where L is the system size, and γ and ν are "critical exponents", which (along with a few others) completely describe the dynamics of the system around the critical point[4].

These critical exponents would allow us to predict the optimal sparsity of our network for any system size.

PetaVision

Using PetaVision, we construct the denoising network seen in Figure 2 and train it by reconstructing 50,000 CIFAR-10 images[3] using varying numbers of features.

We then test the efficiency of the sparse representation found by the V1 layer by measuring the reconstruction error when reconstructing 10,000 CIFAR-10 test images that have had Gaussian noise added to them.

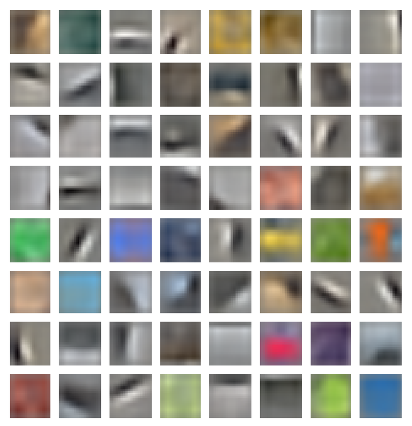


Figure 1: Feature kernel found by the V1 layer for the sparse representation

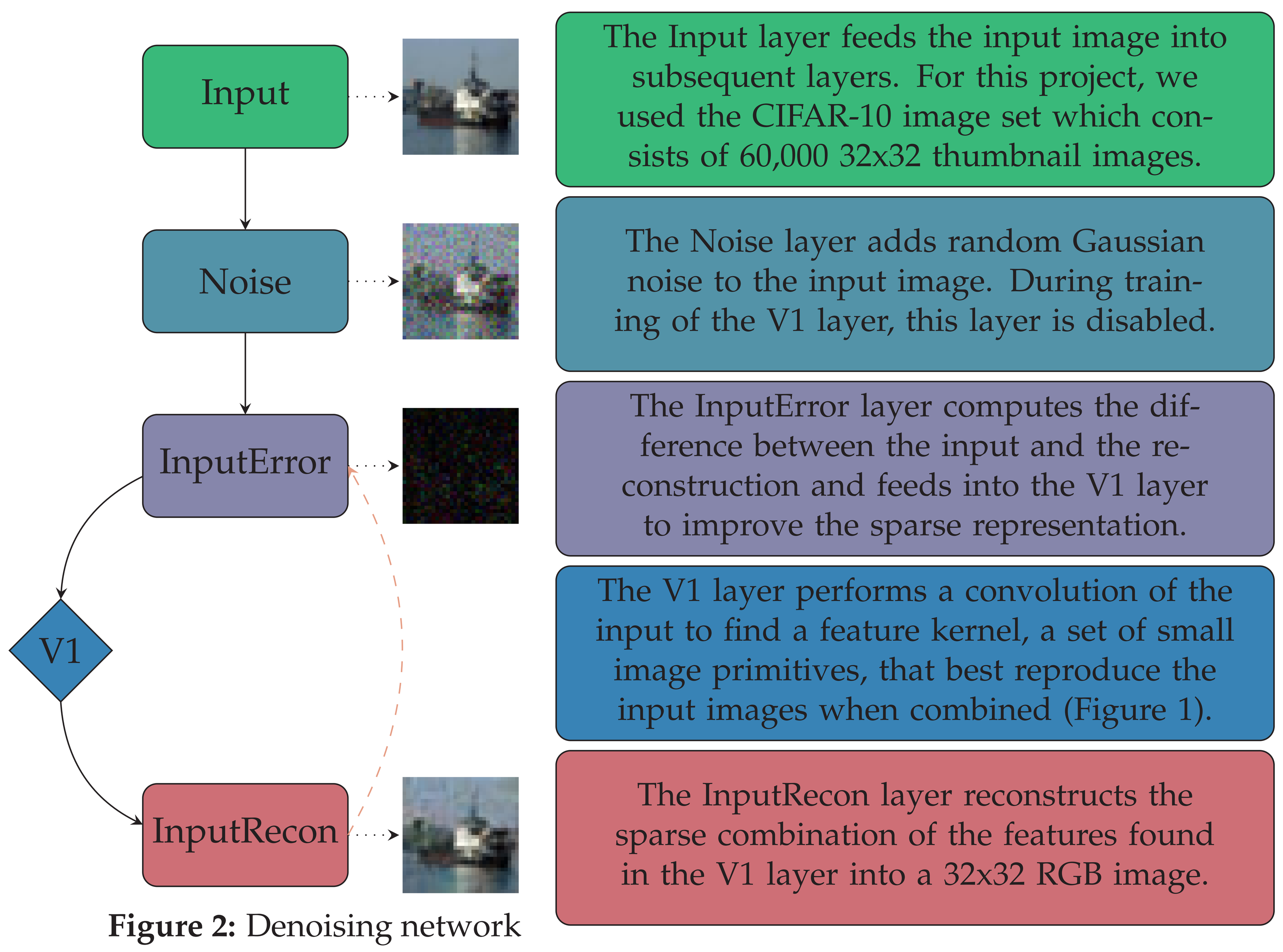
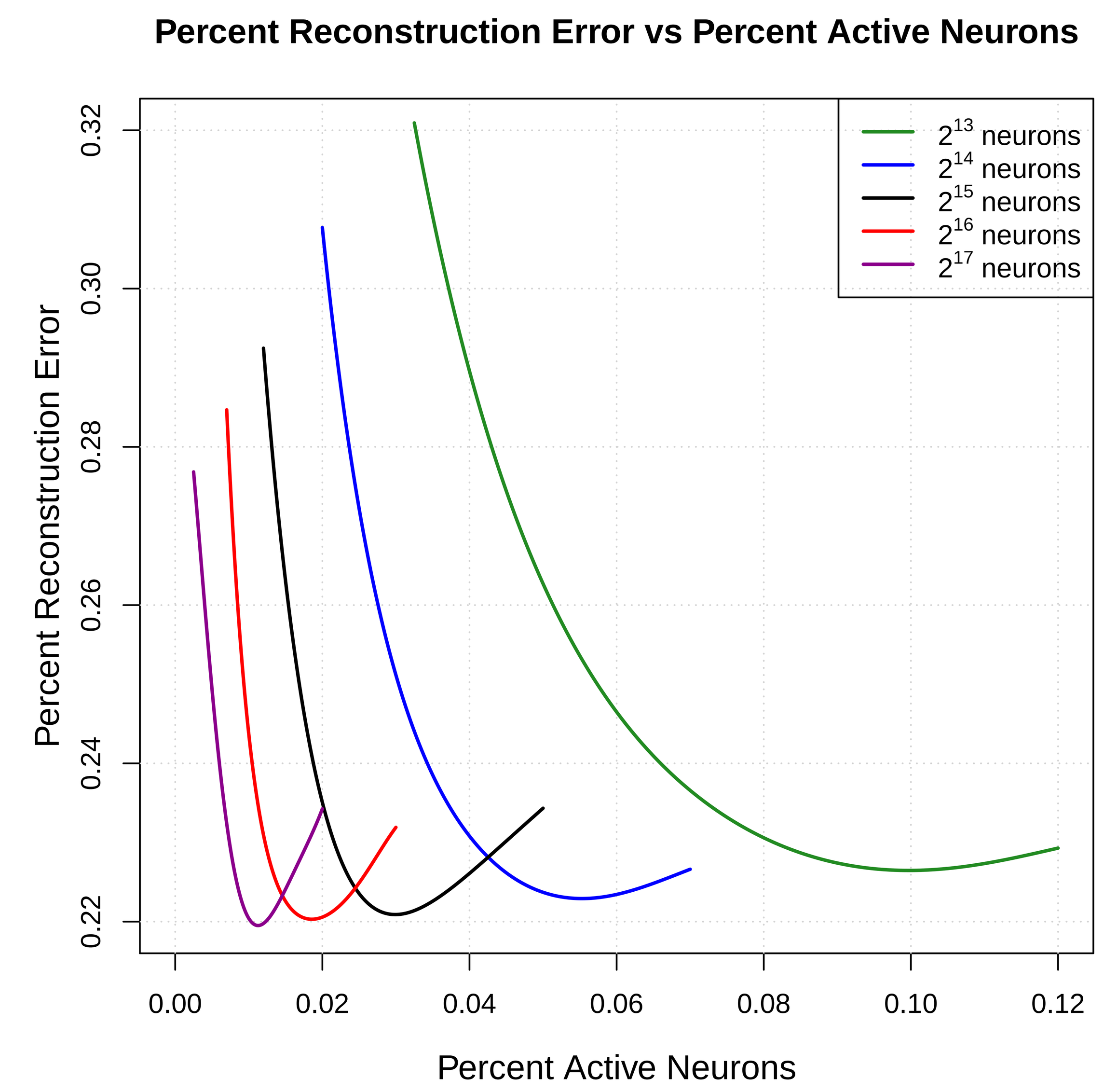
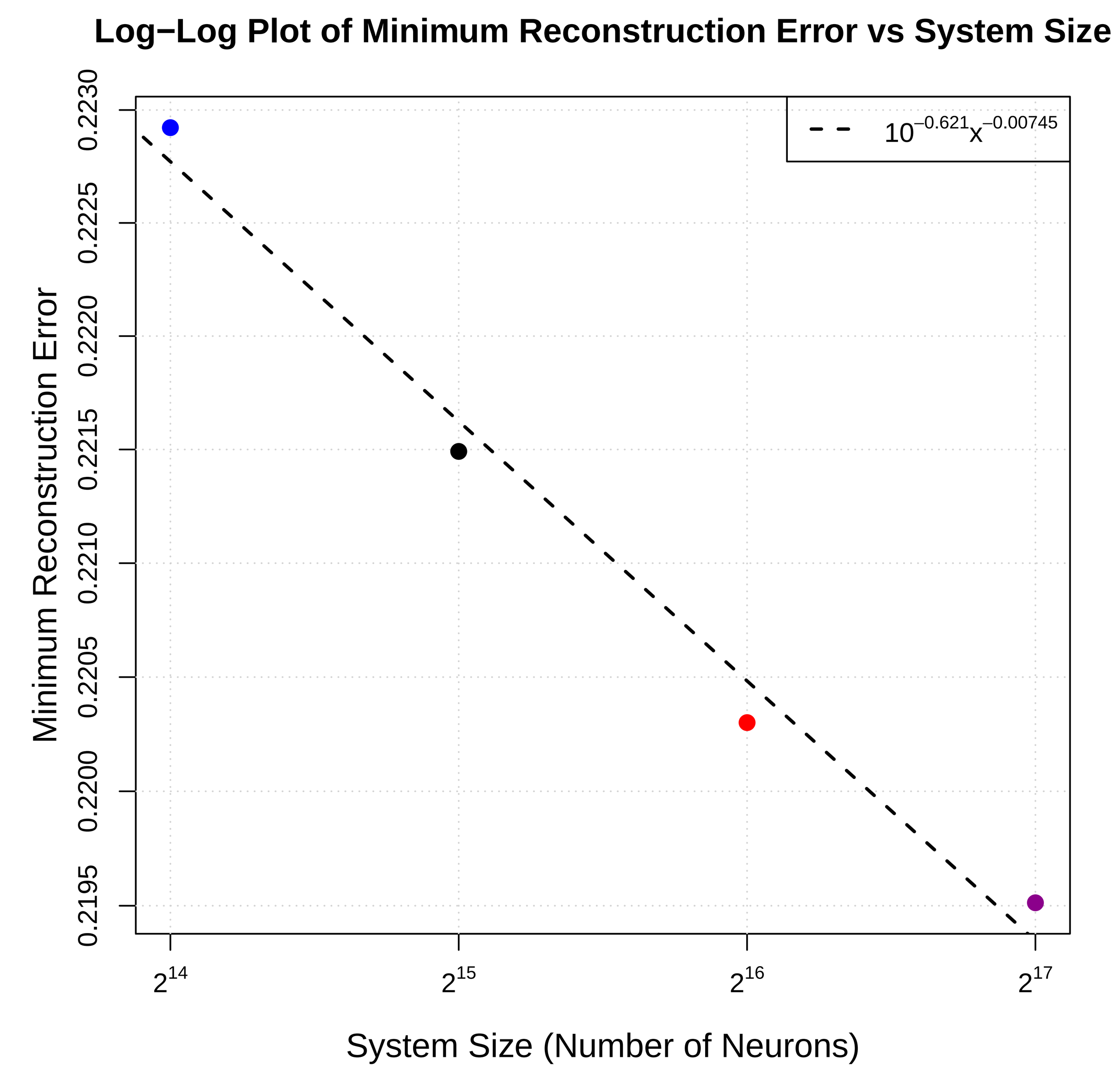
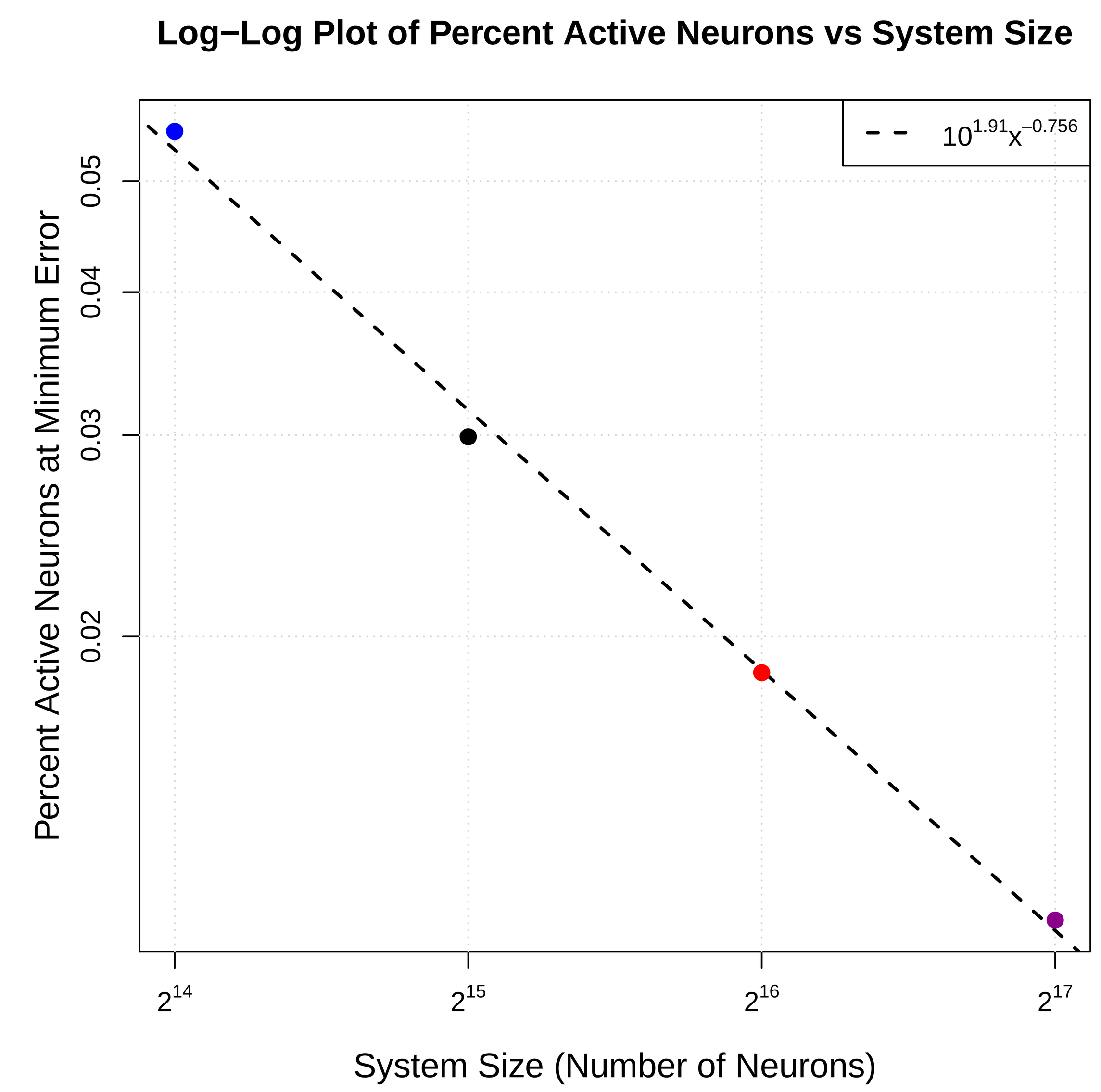


Figure 2: Denoising network

Results







We computed† the reconstruction error vs the percentage of active neurons for many different system sizes (number of neurons). The top left plot shows how the reconstruction error behaves as a function of percent active neurons for different system sizes. The top right plot shows how the minimum reconstruction error behaves as the system size increases. The bottom left plot shows how the percentage of active neurons at the minima changes as system size increases.

Critical Exponent	Value
γ	0.00985
ν	-1.32

Table 1: Extracted critical exponents

We observe the power law scaling behavior required in equations (3) and (4) in the respective plots versus system size. We extract the critical exponents γ and ν via equations (3), (4), and the power law relations found in the system size plots. We report $\gamma = 0.00985$ and $\nu = -1.32$ for this sparsely coded convolutional neural network.

The existence of this power law scaling relation between the minima, percent active and system size verifies that there is a phase transition occuring as the sparsity varies.

† *These results were computed on Power8 nodes consisting of two IBM Power8 CPUs and two Nvidia P100 accelerators. We parallelized the network by doing batch training with 64 MPI ranks per node, and two OpenMP threads per rank.*

Conclusion

The existence of the minima in reconstruction error coupled with the power law behavior of the critical point as the system size is increased demonstrates the existence of a phase transition occurring as the sparsity of the system is varied.

We can improve efficiency of PetaVision by operating at the critical point of this phase transition.

The extracted critical exponents can be used to extrapolate the ideal sparsity for any given system size.

Future Work

Future work for this project entails examining this phase transition in networks constructed for different tasks including image classification, video frame prediction, and compressive sensing.

Additionally, we can quantify the response function χ of the network by measuring how the system responds to perturbations as the image is being denoised. Near the critical point χ should follow the relation $\chi \sim \lambda^{\gamma}$, allowing us to verify our value of γ.

References & Acknowledgements

1. PetaVision. `github.com/PetaVision/OpenPV`
2. Rozell C. J., et al. Sparse Coding via Thresholding and Local Competition in Neural Circuits. *Neural Computation* 20 (2008)
3. Krizhevsky A. Learning Multiple Layers of Features from Tiny Images. (2009)
4. Täuber U. C. *Critical dynamics: a field theory approach to equilibrium and non-equilibrium scaling behavior*. Cambridge University Press (2014)

- This work was supported by the Los Alamos National Laboratory under contract DE-AC52-06NA25396
- Computations were performed using the Darwin Computational Cluster at Los Alamos National Laboratory.
- Thanks to the mentors and organizers of the Parallel Computing Summer Research Internship, Bob Robey, Hai Ah Nam, Kris Garrett, and Joe Schoonover!